Ann. Inst. Statist. Math.
Vol. 57, No. 2, 279–290 (2005)

VARIANCE ESTIMATION FOR SAMPLE QUANTILES USING THE m OUT OF n BOOTSTRAP

K. Y. CHEUNG AND STEPHEN M. S. LEE*

Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong, China, e-mail: celiacky@graduate.hku.hk; smslee@hkusua.hku.hk

(Received January 21, 2004; revised May 7, 2004)

Abstract. We consider the problem of estimating the variance of a sample quantile calculated from a random sample of size n. The r-th-order kernel-smoothed bootstrap estimator is known to yield an impressively small relative error of order $O(n^{-r/(2r+1)})$. It nevertheless requires strong smoothness conditions on the underlying density function, and has a performance very sensitive to the precise choice of the bandwidth. The unsmoothed bootstrap has a poorer relative error of order $O(n^{-1/4})$, but works for less smooth density functions. We investigate a modified form of the bootstrap, known as the m out of n bootstrap, and show that it yields a relative error of order smaller than $O(n^{-1/4})$ under the same smoothness conditions required by the conventional unsmoothed bootstrap on the density function, provided that the bootstrap sample size m is of an appropriate order. The estimator permits exact, simulation-free, computation and has accuracy fairly insensitive to the precise choice of m. A simulation study is reported to provide empirical comparison of the various methods.

Key words and phrases: m out of n bootstrap, quantile, smoothed bootstrap.

1. Introduction

Suppose that $X_1, \ldots, X_n$ constitute a random sample of size n taken from a distribution F. Let $X_{(j)}$ denote the j-th smallest datum in the sample. For a fixed $p \in (0,1)$, assume that F has a continuous and positive density f on $F^{-1}(\mathcal{O})$ for an open neighbourhood $\mathcal{O}$ containing p. Denote by $\xi_p = F^{-1}(p)$ the unique p-th quantile of F. The p-th sample quantile $X_{(r)}$ is a natural and consistent estimator for ξ_p, where $r = [np] + 1$ and $[\cdot]$ denotes the integer part function. Standard theory establishes that $\sigma_n^2 \equiv \mathrm{Var}(X_{(r)})$ admits an asymptotic expansion

$$\sigma_n^2 = n^{-1}p(1-p)f(\xi_p)^{-2} + o(n^{-1}). \tag{1.1}$$

A general discussion can be found in Stuart and Ord ((1994), §10.10). Although (1.1) provides an explicit leading term useful for approximating σ_n^2, its direct computation requires the value of $f(\xi_p)$, which is usually unknown and is difficult to estimate. The conventional, n out of n, unsmoothed bootstrap draws a large number of bootstrap samples, each of size n, from $X_1, \ldots, X_n$, and estimates σ_n^2 by the sample variance of

*Supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. HKU 7131/00P).

the bootstrap sample quantiles calculated from the bootstrap samples. Hall and Martin (1988) show that the theoretical n out of n bootstrap estimator $\hat{\sigma}_n^2$, which is based on infinite simulation of bootstrap samples, has an explicit expression

$$\hat{\sigma}_n^2 = \sum_{j=1}^{n} (X_{(j)} - X_{(r)})^2 w_{n,j}, \tag{1.2}$$

where $w_{n,j} = r\binom{n}{r} \int_{(j-1)/n}^{j/n} x^{r-1}(1-x)^{n-r} dx$. They prove that $\hat{\sigma}_n^2$ has a large relative error of order $O(n^{-1/4})$, that is $\hat{\sigma}_n^2/\sigma_n^2 = 1 + O(n^{-1/4})$. Maritz and Jarrett (1978) note that $\hat{\sigma}_n^2$ may be more accurate than the leading term in the asymptotic formula (1.1) for $p = 1/2$ in small-sample cases, even if the true $f(\xi_p)$ is employed to calculate the latter. The smoothed bootstrap modifies the n out of n bootstrap procedure by drawing (smoothed) bootstrap samples from a kernel density estimate of f rather than from the empirical distribution of $X_1, \ldots, X_n$. Hall *et al.* (1989) show that the smoothed bootstrap estimator has a smaller relative error, of order $O(n^{-r/(2r+1)})$ based on a kernel of order r, under much stronger smoothness conditions on f, provided that the smoothing bandwidth is chosen of order $n^{-1/(2r+1)}$.

The m out of n bootstrap, as pioneered by Bickel and Freedman (1981), provides a method for rectifying bootstrap inconsistency in many nonregular problems: see, for example, Swanepoel (1986) and Athreya (1987). It is, however, generally less efficient than the n out of n bootstrap when the latter is consistent: see, for example, Shao (1994) and Cheung *et al.* (2005). Exceptional cases have been found though. Wang and Taguri (1998) and Lee (1999) improve the n out of n bootstrap by suitably adjusting the resample size m in estimation and confidence interval problems respectively. Arcones (2003) shows that the n out of n bootstrap provides a consistent estimator for the distribution function of sample quantiles with error of order $O(n^{-1/4})$, whilst the m out of n bootstrap reduces the error to order $O(n^{-1/3})$ by use of $m \propto n^{2/3}$. Janssen *et al.* (2001) obtain independently similar results for U-quantiles. We shall show in the present context that the m out of n bootstrap is also effective in reducing the relative error of $\hat{\sigma}_n^2$ under the minimal smoothness conditions same as those required by the n out of n bootstrap on f.

The rest of the paper is organized as follows. Section 2 reviews the smoothed bootstrap method for variance estimation for sample quantiles. Section 3 studies the convergence rate, as well as the asymptotic distribution, of the m out of n bootstrap variance estimator. Section 4 describes a computational algorithm for empirically determining the optimal m. Section 5 presents a simulation study to compare the performances of the various variance estimators. Section 6 concludes our findings. Technical details are given in the Appendix.

2. Smoothed bootstrap

We review the smoothed bootstrap procedure for estimating σ_n^2. Instead of resampling from the empirical distribution of $X_1, \ldots, X_n$, the smoothed bootstrap simulates smoothed bootstrap samples from a kernel density estimate $\hat{f}_b$ of f, given by $\hat{f}_b(x) = (nb)^{-1} \sum_{i=1}^{n} K((x - X_i)/b)$, where $b > 0$ denotes the bandwidth and K is an r-th-order kernel function for $r \geq 2$. The smoothed bootstrap estimator $\hat{\sigma}_{n,b}^2$ of σ_n^2 is then obtained by calculating the sample variance of the p-th-order smoothed bootstrap sample quantiles.

Let $f^{(j)}$ be the j-th-derivative of f. Assume that $f^{(r)}$ exists and is uniformly continuous, $f^{(j)}$ is bounded for $0 \le j \le r$, f is bounded away from 0 in a neighbourhood of ξ_p and $\mathbb{E}|X|^\eta < \infty$ for some $\eta > 0$. Then Hall *et al.* (1989) show that $\hat{\sigma}^2_{n,b}$ has the optimal relative error of order $O(n^{-r/(2r+1)})$, achieved by setting $b \propto n^{-1/(2r+1)}$. In principle, the relative error can be made arbitrarily close to $O(n^{-1/2})$ by choosing a sufficiently high kernel order r.

It should be noted that when $r > 2$, $\hat{f}_b(x)$ necessarily takes on negative values for some x and poses practical difficulties if smoothed bootstrap samples need be simulated from $\hat{f}_b$. Negativity correction techniques of some sort must be incorporated into the smoothed bootstrap procedure to make it computationally feasible: see, for example, Lee and Young (1994). In the case where $r = 2$ so that $\hat{f}_b$ is a proper density function, the optimal relative error of $\hat{\sigma}^2_{n,b}$ is of order $O(n^{-2/5})$, which already improves upon the unsmoothed n out of n bootstrap, which has a relative error of order $O(n^{-1/4})$.

3. m out of n bootstrap

The m out of n bootstrap modifies the n out of n bootstrap by drawing bootstrap samples of size m, instead of n, from the empirical distribution of $X_1, \ldots, X_n$, where m satisfies $m = o(n)$ and $m \to \infty$ as $n \to \infty$. The corresponding variance estimator $\hat{\sigma}^2_m$ is then defined as m/n times the sample variance of the p-th bootstrap sample quantiles.

Recall that $X_{(j)}$ is the j-th order statistic of $X_1, \ldots, X_n$ and $X_{(r)}$ is the p-th sample quantile. The m out of n bootstrap variance estimator $\hat{\sigma}^2_m$ admits an explicit, directly computable, formula:

$$\hat{\sigma}^2_m = (m/n) \sum_{j=1}^{n} (X_{(j)} - X_{(r)})^2 w_{m,j}, \tag{3.1}$$

where $w_{m,j} = k\binom{m}{k} \int_{(j-1)/n}^{j/n} x^{k-1}(1-x)^{m-k} dx$ and $k = [mp]+1$. Our main theorem below establishes asymptotic normality of $\hat{\sigma}^2_m$ together with the corresponding convergence rate. Its proof is outlined in the Appendix.

THEOREM 3.1. *Assume $m \propto n^\lambda$ for some $\lambda \in (0,1)$, $\mathbb{E}|X|^\eta < \infty$ for some $\eta > 0$, $f \equiv F'$ exists and satisfies a Lipschitz condition of order $\nu = \frac{1}{2} + \varepsilon$, with $\varepsilon \in (0, \frac{1}{2}]$, in a neighbourhood of ξ_p, and $f(\xi_p) > 0$. Then*

$$n^{3/2} m^{-1/4} (\hat{\sigma}^2_m - \sigma^2_n) = S_n + O_p(m^{1/4} n^{-1/2} + m^{-1/2-\varepsilon/2} n^{1/2}), \tag{3.2}$$

where S_n converges in distribution to $N(0, 2\pi^{-1/2}[p(1-p)]^{3/2} f(\xi_p)^{-4})$.

The expansion (3.2) enables us to deduce the optimal choice of m by which $\hat{\sigma}^2_m$ achieves the fastest convergence rate, as is asserted in the following corollary.

COROLLARY 3.1. *Under the conditions of Theorem 3.1, $\hat{\sigma}^2_m$ has an optimal relative error of order $O(n^{-(1+2\varepsilon)/(4+4\varepsilon)})$, achieved by setting $m \propto n^{1/(1+\varepsilon)}$.*

Hall and Martin (1988) show that $n^{5/4}(\hat{\sigma}^2_n - \sigma^2_n)$ has the same asymptotic normal distribution as does $n^{3/2} m^{-1/4} (\hat{\sigma}^2_m - \sigma^2_n)$ under exactly the same conditions of Theorem 3.1. It is clear that $\hat{\sigma}^2_m$ converges to σ^2_n at a faster rate than does $\hat{\sigma}^2_n$, which has a relative error of order $O(n^{-1/4})$. Although the smoothed bootstrap estimator $\hat{\sigma}^2_{n,b}$ has an even smaller relative error, of order $O(n^{-r/(2r+1)})$, than $\hat{\sigma}^2_m$ for any $r \geq 2$, it requires that f be at least twice continuously differentiable in a neighbourhood of ξ_p, a condition much stronger than those of Theorem 3.1. Moreover, that no such computable expression as (3.1) exists for $\hat{\sigma}^2_{n,b}$ means that $\hat{\sigma}^2_{n,b}$ has to be approximated by Monte Carlo simulation, which is computationally more expensive and is not immediately feasible if $r > 2$ due to the problem of negativity of $\hat{f}_b$.

Arcones (2003) establishes versions of Theorem 3.1 and Corollary 3.1 for m out of n bootstrap estimation of the distribution of $X_{(r)}$. He shows, under the stronger assumption that f is differentiable at ξ_p, that the fastest convergence rate, of order $n^{-1/3}$, is attained by setting $m \propto n^{2/3}$. Our results apply to the variance of $X_{(r)}$ and to densities f under less stringent smoothness conditions. Densities violating Arcones' but satisfying our smoothness conditions include those which are Lipschitz continuous of order $\nu \in (1/2, 1)$ near ξ_p. A simple example is $f(x) = (7/6)(1 - |x|^{3/4})$, for $|x| \leq 1$, which is Lipschitz continuous of order 3/4 at $x = 0$.

4. Empirical determination of m

It follows from Corollary 3.1 that fixing $m = cn^{\gamma}$, for some constants c and γ independent of n, yields the best convergence rate for $\hat{\sigma}^2_m$. In practice γ is unknown and so is the optimal value of c. We sketch below a simple algorithm, based on the bootstrap, for empirical determination of both c and γ and hence the optimal choice of m.

First fix S distinct bootstrap sample sizes $m_1, \ldots, m_S < n$, for some $S \geq 2$. For each $s = 1, \ldots, S$, calculate $\sigma^{*2}_s = (n/m_s)\hat{\sigma}^2_{m_s}$, the variance of the p-th bootstrap sample quantile induced by the drawing of bootstrap samples of size m_s. Generate a large number, B say, of bootstrap samples $\mathcal{X}^*_{s,1}, \ldots, \mathcal{X}^*_{s,B}$, each of size m_s, from $X_1, \ldots, X_n$. For each $\mathcal{X}^*_{s,b}$, calculate the ℓ out of m_s estimate of σ^{*2}_s, namely

$$\hat{\sigma}^{*2}_{s,b,\ell} = (\ell/m_s) \sum_{j=1}^{m_s} (X^*_{b,(j)} - X^*_{b,(r^*)})^2 k^* \binom{\ell}{k^*} \int_{(j-1)/m_s}^{j/m_s} x^{k^*-1}(1-x)^{\ell-k^*}\, dx,$$

where $k^* = [\ell p] + 1$, $r^* = [m_s p] + 1$ and $X^*_{b,(i)}$ denotes the i-th smallest datum in $\mathcal{X}^*_{s,b}$. The mean squared error of the ℓ out of m_s bootstrap variance estimate is then estimated by $MSE_s(\ell) = B^{-1}\sum_{b=1}^{B}(\hat{\sigma}^{*2}_{s,b,\ell} - \sigma^{*2}_s)^2$. Select $\ell = \ell_s$ which minimizes $MSE_s(\ell)$ over $\ell \in \{1, \ldots, m_s\}$. Asymptotically $\ell_s \approx cm_s^{\gamma}$, so that $\log \ell_s \approx \log c + \gamma \log m_s$, for $s = 1, \ldots, S$. Standard least squares techniques yield that $c \approx \exp\{D^{-1}(M_2 L_1 - M_1 K)\}$ and $\gamma \approx D^{-1}(SK - M_1 L_1)$, where $M_1 = \sum_{s=1}^{S} \log m_s$, $M_2 = \sum_{s=1}^{S} (\log m_s)^2$, $L_1 = \sum_{s=1}^{S} \log \ell_s$, $K = \sum_{s=1}^{S} (\log m_s)(\log \ell_s)$ and $D = SM_2 - M_1^2$. Finally calculate the optimal m to be $m = [cn^{\gamma}]$, with c and γ fixed at the above approximate values.

5. Simulation study

We conducted a simulation study to compare the mean squared errors of $\hat{\sigma}^2_n$, $\hat{\sigma}^2_m$ and $\hat{\sigma}^2_{n,b}$, for $p = 0.1$, 0.5 and 0.9 and for fixed values of m and b. Random samples of sizes $n =$ 50 and 200 were generated from three distributions: (i) the standard normal distribution, $N(0,1)$, (ii) the chi-squared distribution with 5 degrees of freedom, χ^2_5, and (iii) the double exponential distribution with density function $f(x) = (1/2)\exp(-|x|)$. All three distributions have densities satisfying the Lipschitz condition of order one, so that the

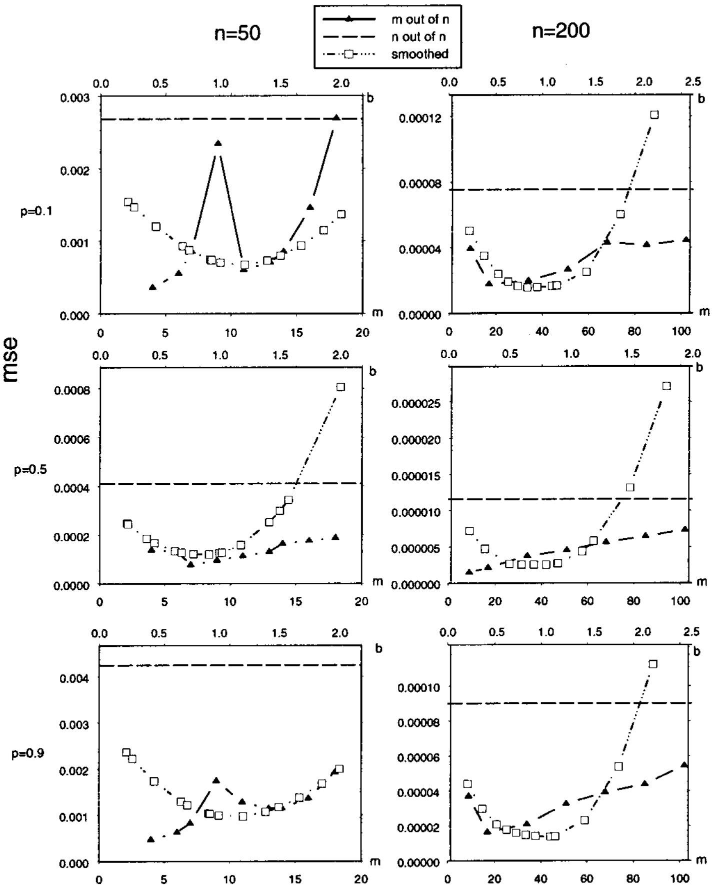

Fig. 1. Normal Example: mean squared errors of $\hat{\sigma}_n^2$, $\hat{\sigma}_m^2$ (plotted against m) and $\hat{\sigma}_{n,b}^2$ (plotted against b) for $n = 50$ and 200, and $p = 0.1$, 0.5 and 0.9.

conditions of Theorem 3.1 hold for $\varepsilon = 1/2$. For the smoothed bootstrap estimator $\hat{\sigma}_{n,b}^2$, the second-order Epanechnikov kernel function $k(t) = \max\{(3/4)(1 - t^2), 0\}$ was employed. Note that the first derivative of the double exponential density function does not exist at $\xi_{0.5} = 0$, so that the density there lacks the smoothness condition sufficient for proper functioning of the smoothed bootstrap method based on the kernel k above. Each smoothed bootstrap estimate $\hat{\sigma}_{n,b}^2$ was derived from 1,000 smoothed bootstrap samples. The estimates $\hat{\sigma}_n^2$ and $\hat{\sigma}_m^2$ were directly computed using explicit formulae (1.2) and (3.1) respectively. Each mean squared error was obtained by averaging over 1,600 random samples drawn from F.

Figure 1 plots the mean squared error of $\hat{\sigma}_m^2$ against m (bottom axis) and that of $\hat{\sigma}_{n,b}^2$ against b (top axis) for the normal distribution. Similar comparisons for the chi-squared and double exponential distributions are given in Figs. 2 and 3 respectively. The mean squared error of $\hat{\sigma}_n^2$ is also included in each diagram for reference.

For the $N(0, 1)$ data, as predicted from asymptotic results, the n out of n bootstrap yields for $\hat{\sigma}_n^2$ the largest mean squared error, except for cases of large b or m, for all combinations of n and p. The mean squared error of the smoothed bootstrap estimate

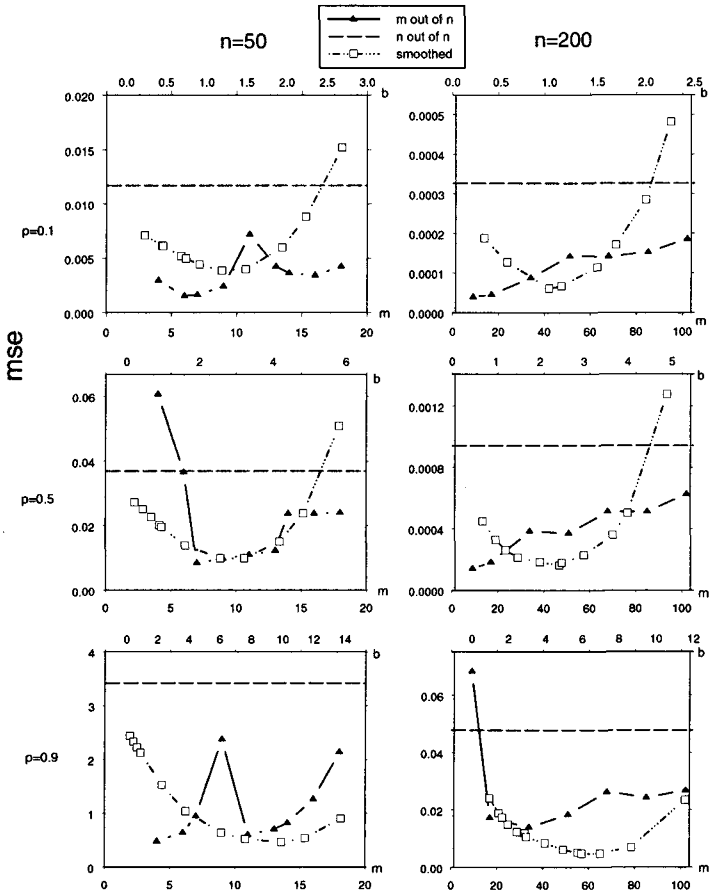

Fig. 2. Chi-squared Example: mean squared errors of $\hat{\sigma}_n^2$, $\hat{\sigma}_m^2$ (plotted against m) and $\hat{\sigma}_{n,b}^2$ (plotted against b) for $n = 50$ and 200, and $p = 0.1$, 0.5 and 0.9.

$\hat{\sigma}_{n,b}^2$ varies with b parabolically. Although it is asymptotically less accurate, the m out of n bootstrap estimate $\hat{\sigma}_m^2$ has mean squared error comparable to that of $\hat{\sigma}_{n,b}^2$ constructed using an optimal b, and maintains a more stable performance than $\hat{\sigma}_{n,b}^2$ for $n = 200$. Among the values of p studied, all three estimators tend to be most accurate at $p = 0.5$ for both $n = 50$ and 200.

For data drawn from the asymmetric χ_5^2, the mean squared errors of the estimators are in general larger than those observed in the $N(0,1)$ example, and increase as p increases. As in Fig. 1, we see from Fig. 2 that $\hat{\sigma}_n^2$ is generally the least accurate, while the mean squared errors of $\hat{\sigma}_m^2$ and $\hat{\sigma}_{n,b}^2$ are of similar magnitudes. The optimal choice of bandwidth, which yields the minimum mean squared error for $\hat{\sigma}_{n,b}^2$, increases considerably as p increases; the optimal choice of m for $\hat{\sigma}_m^2$, by constrast, stays within the same range as p varies, rendering its empirical determination less difficult than that of the optimal bandwidth.

Figure 3 displays the findings for the double exponential data. For $p = 0.1$ and 0.9, we see that $\hat{\sigma}_{n,b}^2$ and $\hat{\sigma}_m^2$ have comparable mean squared errors, which are notably smaller than that of $\hat{\sigma}_n^2$, provided b and m are selected sensibly. For $p = 0.5$, the mean squared

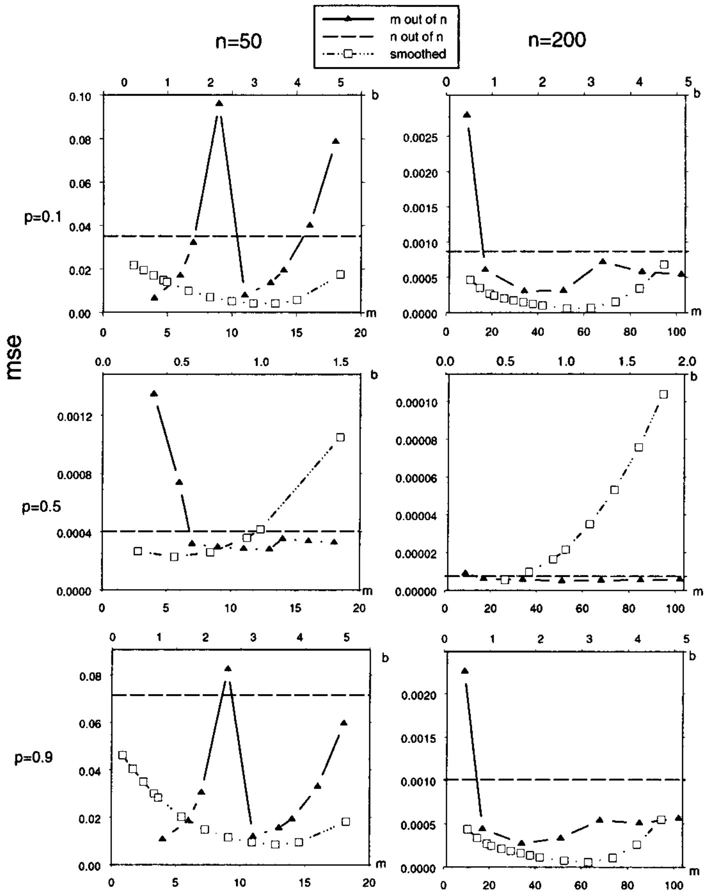

Fig. 3. Double Exponential Example: mean squared errors of $\hat{\sigma}_n^2$, $\hat{\sigma}_m^2$ (plotted against m) and $\hat{\sigma}_{n,b}^2$ (plotted against b) for $n = 50$ and 200, and $p = 0.1$, 0.5 and 0.9.

error of $\hat{\sigma}_{n,b}^2$ increases significantly as b increases, and is much larger than those of $\hat{\sigma}_n^2$ and $\hat{\sigma}_m^2$ for $n = 200$, due plausibly to the lack of smoothness of the double exponential density at $\xi_{0.5} = 0$. In general the m out of n bootstrap performs much better than the n out of n bootstrap for $n = 200$, except for a small $m = 9$. Similar to the $N(0,1)$ example, all three methods are most accurate at $p = 0.5$ among the values of p studied.

We note that in most of the investigated cases the accuracy of the m out of n bootstrap deteriorates markedly for some very small values of m. A heuristic explanation is as follows. We see from the proof of Theorem 3.1 that asymptotic properties of $\hat{\sigma}_m^2$ depend critically on the weights $w_{m,j}$ for j close to r. Lemma A.1 shows that the $w_{m,j}$ sequence, for j close to r, resembles asymptotically the central shape of a normal density. Thus our asymptotic findings can reliably predict finite-sample behaviour only when the $w_{m,j}$ attains its mode at some j strictly between 1 and n. Examination of the $w_{m,j}$ in detail shows that the latter condition holds only when m exceeds a certain value, $M(n,p)$ say, depending on both n and p. Under the settings of our simulation study, we find that for both $n = 50$ and 200, $M(n,p) = 9$, 2, 10 for $p = 0.1$, 0.5 and 0.9 respectively. Indeed Figs. 1–3 all suggest that the m out of n bootstrap performance begins to stabilize once

Table 1. Mean squared errors of various variance estimates. In the case of $\hat{\sigma}^2_m$, results are shown for both the smallest error obtained in the simulation study using fixed m and the error given by empirically selecting m using the algorithm in Section 4. Mean and standard deviation of the empirical m are also included.

Normal example

	$n = 50$			$n = 200$		
	$p = 0.1$	$p = 0.5$	$p = 0.9$	$p = 0.1$	$p = 0.5$	$p = 0.9$
$\hat{\sigma}^2_n$	2.7×10^{-3}	4.1×10^{-4}	4.2×10^{-3}	7.5×10^{-5}	1.2×10^{-5}	9.0×10^{-5}
$\hat{\sigma}^2_{n,b}$ (fixed b)	6.7×10^{-4}	1.2×10^{-4}	9.8×10^{-4}	1.6×10^{-5}	2.5×10^{-6}	1.4×10^{-5}
$\hat{\sigma}^2_m$ (fixed m)	6.0×10^{-4}	7.8×10^{-5}	1.1×10^{-3}	1.8×10^{-5}	1.5×10^{-6}	1.6×10^{-5}
$\hat{\sigma}^2_m$ (empirical m)	1.5×10^{-3}	1.2×10^{-4}	2.0×10^{-3}	2.3×10^{-5}	3.0×10^{-6}	3.5×10^{-5}
mean of empirical m	*8.0*	*5.9*	*7.9*	*11.5*	*7.8*	*13.5*
s.d. of empirical m	*6.5*	*4.2*	*6.2*	*11.3*	*8.8*	*14.5*

Chi-squared example

	$n = 50$			$n = 200$		
	$p = 0.1$	$p = 0.5$	$p = 0.9$	$p = 0.1$	$p = 0.5$	$p = 0.9$
$\hat{\sigma}^2_n$	1.2×10^{-2}	3.7×10^{-2}	3.4×10^{0}	3.3×10^{-4}	9.4×10^{-4}	4.8×10^{-2}
$\hat{\sigma}^2_{n,b}$ (fixed b)	3.9×10^{-3}	9.8×10^{-3}	4.7×10^{-1}	6.1×10^{-5}	1.6×10^{-4}	4.8×10^{-2}
$\hat{\sigma}^2_m$ (fixed m)	3.4×10^{-3}	8.4×10^{-3}	6.0×10^{-1}	4.5×10^{-5}	1.4×10^{-4}	1.4×10^{-2}
$\hat{\sigma}^2_m$ (empirical m)	3.5×10^{-3}	1.4×10^{-2}	1.4×10^{0}	1.0×10^{-4}	3.3×10^{-4}	2.5×10^{-2}
mean of empirical m	*8.3*	*7.7*	*10.1*	*8.2*	*7.6*	*13.9*
s.d. of empirical m	*4.2*	*8.9*	*7.1*	*5.8*	*9.4*	*12.3*

Double exponential example

	$n = 50$			$n = 200$		
	$p = 0.1$	$p = 0.5$	$p = 0.9$	$p = 0.1$	$p = 0.5$	$p = 0.9$
$\hat{\sigma}^2_n$	3.5×10^{-2}	4.0×10^{-4}	7.2×10^{-2}	8.6×10^{-4}	7.4×10^{-6}	1.0×10^{-3}
$\hat{\sigma}^2_{n,b}$ (fixed b)	4.1×10^{-3}	2.3×10^{-4}	8.7×10^{-3}	5.9×10^{-5}	5.4×10^{-6}	5.7×10^{-5}
$\hat{\sigma}^2_m$ (fixed m)	7.8×10^{-3}	2.8×10^{-4}	1.2×10^{-2}	3.0×10^{-4}	4.8×10^{-6}	2.7×10^{-4}
$\hat{\sigma}^2_m$ (empirical m)	2.9×10^{-2}	3.5×10^{-4}	3.2×10^{-2}	3.9×10^{-4}	8.3×10^{-6}	6.3×10^{-4}
mean of empirical m	*9.1*	*15.7*	*10.3*	*13.1*	*34.6*	*14.8*
s.d. of empirical m	*7.8*	*19.3*	*7.3*	*11.1*	*34.2*	*13.5*

m exceeds $M(n,p)$, especially for $n = 200$. On the other hand, the optimal choice of bandwidth for $\hat{\sigma}^2_{n,b}$ depends crucially on F, n and p, and its mean squared error increases considerably if b deviates from its optimal value.

Table 1 compares numerically the mean squared error of $\hat{\sigma}^2_n$ with those of $\hat{\sigma}^2_m$ and $\hat{\sigma}^2_{n,b}$ at the optimal choices of m (among values greater than $M(n,p)$) and b as observed from the simulation study. In the case of $\hat{\sigma}^2_m$, we include also results obtained using m selected by the algorithm described in Section 4, in which 1,000 pilot bootstrap samples were simulated to estimate the mean squared error of the ℓ out of m_s bootstrap variance estimate and the m_s were chosen to be $2s + 8$ for $n = 50$ and $12s - 2$ for $n = 200$, $s = 1, \ldots, 8$. The mean and standard deviation of the empirical choice of m are reported alongside the mean squared error findings. We see that the optimally constructed $\hat{\sigma}^2_m$ and $\hat{\sigma}^2_{n,b}$, at fixed m and b respectively, have comparable errors. Both of them are considerably more accurate than $\hat{\sigma}^2_n$. In general, our algorithm for empirical determination of m worked satisfactorily and produced estimates more accurate than $\hat{\sigma}^2_n$, albeit to a lesser extent than its fixed-m counterpart.

6. Conclusion

We have shown, both theoretically and empirically, that the m out of n bootstrap variance estimator $\hat{\sigma}_m^2$ is notably superior to the conventional n out of n bootstrap estimator $\hat{\sigma}_n^2$. For densities satisfying a Lipschitz condition of order within $(1/2, 1]$ near ξ_p, $\hat{\sigma}_m^2$ incurs a relative error of smaller order than $\hat{\sigma}_n^2$, provided that m is chosen appropriately. The smoothed bootstrap estimator $\hat{\sigma}_{n,b}^2$ may yield an even smaller relative error using an optimal bandwidth b, but requires much stronger smoothness conditions on the density f. The m out of n bootstrap therefore offers a convenient alternative which is more accurate than the n out of n bootstrap and more robust than the smoothed bootstrap. Under a smooth f for which both smoothed and unsmoothed bootstraps work properly, we have that $\hat{\sigma}_n^2$, $\hat{\sigma}_m^2$ and $\hat{\sigma}_{n,b}^2$ generate relative errors of orders $O(n^{-1/4})$, $O(n^{-1/3})$ and $O(n^{-2/5})$ respectively, provided that $m \propto n^{2/3}$, $b \propto n^{-1/5}$ and a second-order kernel is used in constructing $\hat{\sigma}_{n,b}^2$.

Our simulation results agree closely with the asymptotic findings. Both the smoothed and the m out of n bootstraps, when constructed optimally, yield comparable accuracies and outperform the n out of n bootstrap method substantially. The optimal choice of bandwidth for the smoothed bootstrap varies considerably with the problem settting. The mean squared error of $\hat{\sigma}_{n,b}^2$ is also very sensitive to the bandwidth. A slight deviation from the optimal value of the bandwidth may greatly deteriorate the accuracy of the estimate. One therefore requires a sophisticatedly-designed, data-dependent, procedure for calculating the optimal bandwidth in practice. On the other hand, the observed mean squared error of $\hat{\sigma}_m^2$ remains relatively stable over a wide range of m beyond $M(n,p)$, especially for large n. Also, the optimal choice of m tends to stay within a stable region which varies little with the problem setting. This suggests that the precise determination of m is less crucial an issue than is the choice of bandwidth for $\hat{\sigma}_{n,b}^2$. We have proposed a simple bootstrap-based algorithm for empirically determining the optimal m and obtained satisfactory results in our simulation study.

Unlike most bootstrap-based estimates, $\hat{\sigma}_n^2$ and $\hat{\sigma}_m^2$ can be evaluated directly using formulae (1.2) and (3.1) respectively, so that no Monte Carlo simulation is necessary, making their computation exact and very efficient. The smoothed bootstrap estimate $\hat{\sigma}_{n,b}^2$ must, however, most conveniently be approximated using Monte Carlo simulation. Use of a higher-order kernel, which effects in an improved error rate, further complicates the Monte Carlo procedure due to negativity of the kernel estimate $\hat{f}_b$.

Appendix

A.1 *Proof of Theorem* 3.1

The proof is modelled after Hall and Martin's (1988) arguments.

Let ϕ denote the standard normal density function, $y_{n,j} = (j-1)/n$ and $b_{mn} = (my_{n,j} - k)\{my_{n,j}(1-y_{n,j})\}^{-1/2}$. The following lemma states a useful asymptotic expansion for the weight $w_{m,j}$.

LEMMA A.1. *Assume that* $m \propto n^{\lambda}$ *for some* $\lambda \in (0,1)$. *There exists some constant* $C > 0$ *such that*

$$w_{m,j} = m^{1/2} n^{-1} \{y_{n,j}(1-y_{n,j})\}^{-1/2} \phi(b_{mn}) + O(n^{-1} e^{-Cm(y_{n,j}-p)^2}).$$

PROOF. Note that $w_{m,j} = I_{j/n}(k, m-k+1) - I_{(j-1)/n}(k, m-k+1)$, where $I_y(a,b) = \sum_{j=a}^{a+b-1} \binom{a+b-1}{j} y^j (1-y)^{a+b-1-j}$. Without loss of generality, consider $j = np + q$ with $q \geq 0$. The proof is completed by considering the Edgeworth expansion of the binomial distribution function for the case $0 \leq q \leq Dnm^{-1/2}(\ln m)^{1/2}$, for some $D > 0$, and Bernstein's inequality for the case $q > Dnm^{-1/2}(\ln m)^{1/2}$. □

We first consider the summation over j in (3.1). The expansion for $\hat{\sigma}_m^2$ then follows trivially after multiplication by m/n. The summation is divided into two parts, for some $\delta > 0$ and $\beta < \lambda/12$: (i) $|j-r| > \delta n^{1+\beta} m^{-1/2}$; and (ii) $|j-r| \leq \delta n^{1+\beta} m^{-1/2}$.

For part (i), we note that $\max\{(X_{(j)} - X_{(r)})^2 : j \leq n\} \leq 4n^{4/\eta}$ in probability: see Hall and Martin (1988). Lemma A.1 implies that, for some constant $C_2 > 0$, $w_{m,j} < C_2 m^{1/2} n^{-1} e^{-Cm(y_{n,j}-p)^2}$. Thus, with probability tending to one, we have that for some constant $C_3 > 0$ and any $\zeta > 0$,

$$\sum_{|j-r|>\delta n^{1+\beta}m^{-1/2}} (X_{(j)} - X_{(r)})^2 w_{m,j} \leq 4\, C_2 m^{1/2} n^{4/\eta} e^{-C_3 n^{2\beta}} = O(n^{-\zeta}). \tag{A.1}$$

For part (ii), we assume throughout that $|j-r| \leq \delta n^{1+\beta} m^{-1/2}$, and that $\sum_j$ refers to summation over j satisfying the above, unless specified otherwise. Let $H(x) = F^{-1}(e^{-x})$ and $Y_1, \ldots, Y_n$ denote independent and identically distributed exponential variables with unit mean. Define $s_j = \mathrm{sgn}(r-j)$, $m_{0j} = \min(r,j)$, $m_{1j} = \max(r,j) - 1$, $A_r = \sum_{u=r}^n u^{-1}$. Suppose that f satisfies a Lipschitz condition of order $\nu = \frac{1}{2} + \varepsilon$ in a neighbourhood of ξ_p, so that $a \equiv H'(A_r) = -pf(\xi_p)^{-1} + O(n^{-1})$. Following Hall and Martin's (1988) arguments, we have

$$\sum_j (X_{(j)} - X_{(r)})^2 w_{m,j} = S_1 + S_2 + T_1 + T_2 + T_3, \tag{A.2}$$

where $S_1 = a^2 \sum_j b_j^2 w_{m,j}$, $S_2 = 2a^2 \sum_j b_j(B_j - b_j) w_{m,j}$, $T_1 = a^2 \sum_j (B_j - b_j)^2 w_{m,j}$, $T_2 = 2\sum_j D_j R_{1j} w_{m,j}$, $T_3 = \sum_j R_{1j}^2 w_{m,j}$, $B_j = \sum_{u=m_{0j}}^{m_{1j}} u^{-1} Y_u$, $b_j = \mathbb{E}(B_j)$, $D_j = s_j a B_j$, $R_{1j} = R_{2j} + R_{3j}$, $R_{2j} = s_j B_j [H'(A_r) - a]$ and $R_{3j} = s_j B_j \int_0^1 [H'(A_r + t s_j B_j) - H'(A_r)] dt$. Note also that $B_r = b_r = 0$ and that

$$b_j = |j-r| r^{-1} + 2^{-1}(j-r)^2 r^{-2} + O(|j-r|^3 r^{-3}). \tag{A.3}$$

Using Lemma A.1 and (A.3), we have

$$\begin{aligned} \sum_j b_j^2 w_{m,j} &= m^{1/2} n^{-1} \sum_j \left(\frac{j-r}{np}\right)^2 \frac{1}{\sqrt{p(1-p)}} \phi\left(\frac{m^{1/2}(j-r-1)}{n\sqrt{p(1-p)}}\right) \\ &\quad + O\left(m^{1/2} n^{-4} \sum_j |j-r|^3 \phi\left(\frac{m^{1/2}(j-r-1)}{n\sqrt{p(1-p)}}\right)\right) \\ &= m^{-1} p^{-1} (1-p) + O(m^{-3/2}), \end{aligned}$$

so that

$$S_1 = m^{-1} p(1-p) f(\xi_p)^{-2} + O(m^{-3/2}). \tag{A.4}$$

Consider next

$$S_2 = 2a^2\left\{ \sum_{u=r-\delta n^{1+\beta}m^{-1/2}}^{r-1} u^{-1}(Y_u - 1) \sum_{j=r-\delta n^{1+\beta}m^{-1/2}}^{u} b_j w_{m,j} + \sum_{u=r}^{r+\delta n^{1+\beta}m^{-1/2}-1} u^{-1}(Y_u - 1) \sum_{j=u+1}^{r+\delta n^{1+\beta}m^{-1/2}} b_j w_{m,j} \right\},$$

so that, by Lyapounov's central limit theorem,

$$m^{3/4}n^{1/2}S_2 \xrightarrow{\mathcal{D}} N(0, 2\pi^{-1/2}[p(1-p)]^{3/2}f(\xi_p)^{-4}). \tag{A.5}$$

We note, using Lemma A.1 again, that for any $t > 0$,

$$\begin{aligned} &\sum_j |j - np|^t w_{m,j} \\ &\sim m^{1/2}n^t \int_{p-\delta n^\beta m^{-1/2}}^{p+\delta n^\beta m^{-1/2}} |y-p|^t [y(1-y)]^{-1/2}\phi\left(\frac{m^{1/2}(y-p)}{\sqrt{y(1-y)}}\right) dy \\ &= O(m^{-t/2}n^t). \end{aligned} \tag{A.6}$$

It follows by substituting appropriate values for t in (A.6) that

$$\mathbb{E}(T_1) = O\left(\sum_j |j - r| r^{-2} w_{m,j}\right) = O(m^{-1/2}n^{-1}),$$

$$\begin{aligned} \mathbb{E}|T_2| &= O\left(\sum_j [n^{-2}(j-r)^2 n^{-1/4-\varepsilon/2} + (n^{-1}|j-r|)^{5/2+\varepsilon}] w_{m,j}\right) \\ &= O(m^{-5/4-\varepsilon/2}) \end{aligned}$$

and

$$\mathbb{E}(T_3) = O\left(\sum_j [n^{-2}(j-r)^2 n^{-1/2-\varepsilon} + (n^{-1}|j-r|)^{3+2\varepsilon}] w_{m,j}\right) = O(m^{-3/2-\varepsilon}),$$

so that, by Chebyshev's inequality,

$$T_1 = O_p(m^{-1/2}n^{-1}), \qquad T_2 = O_p(m^{-5/4-\varepsilon/2}), \qquad T_3 = O_p(m^{-3/2-\varepsilon}). \tag{A.7}$$

Recall, by Hall and Martin's (1988) Theorem 2.1, that

$$\sigma_n^2 = n^{-1}p(1-p)f(\xi_p)^{-2} + O(n^{-3/2-\varepsilon}). \tag{A.8}$$

Subtracting (A.8) from $\hat{\sigma}_m^2$, and expanding the summation in (3.1) using (A.1), (A.2), (A.4), (A.5) and (A.7), we prove (3.2).

A.2 *Proof of Corollary* 3.1

Note that $m \propto n^\lambda$. It follows from (3.2) that the optimal value of λ is obtained by minimizing $\max\{\lambda/4 - 3/2, -\lambda(1/4 + \varepsilon/2) - 1\}$ over $\lambda \in (0,1)$. Corollary 3.1 then follows by using standard linear programming to obtain the optimal λ.

References

Arcones, M. A. (2003). On the asymptotic accuracy of the bootstrap under arbitrary resampling size, *Annals of the Institute of Statistical Mathematics*, **55**, 563–583.

Athreya, K. B. (1987). Bootstrap of the mean in the infinite variance case, *Annals of Statistics*, **15**, 724–731.

Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap, *Annals of Statistics*, **9**, 1196–1217.

Cheung, K. Y., Lee, S. M. S. and Young, G. A. (2005). Stein confidence sets based on non-iterated and iterated parametric bootstraps, *Statistica Sinica* (to appear).

Hall, P. and Martin, M. A. (1988). Exact convergence rate of bootstrap quantile variance estimator, *Probability Theory and Related Fields*, **80**, 261–268.

Hall, P., DiCiccio, T. J. and Romano, R. (1989). On smoothing and the bootstrap, *Annals of Statistics*, **17**, 692–704.

Janssen, P., Swanepoel, J. and Veraverbeke, N. (2001). Modified bootstrap consistency rates for U-quantiles, *Statistics and Probability Letters*, **54**, 261–268.

Lee, S. M. S. (1999). On a class of m out of n bootstrap confidence intervals, *Journal of the Royal Statistical Society. Series B*, **61**, 901–911.

Lee, S. M. S. and Young, G. A. (1994). Practical higher-order smoothing of the bootstrap, *Statistica Sinica*, **4**, 445–459.

Maritz, J. S. and Jarrett, R. G. (1978). A note on estimating the variance of the sample median, *Journal of the American Statistical Association*, **73**, 194–196.

Shao, J. (1994). Bootstrap sample size in nonregular cases, *Proceedings of the American Mathematical Society*, **122**, 1251–1262.

Stuart, A. and Ord, J. K. (1994). *Kendall's Advanced Theory of Statistics*, Edward Arnold, New York.

Swanepoel, J. W. H. (1986). A note on proving that the (modified) bootstrap works, *Communications in Statistics A—Theory and Methods*, **15**, 3193–3203.

Wang, J. and Taguri, M. (1998). Improved bootstrap through modified resample size, *Journal of the Japan Statistical Society*, **28**, 181–192.